

DOCUMENT RESUME

ED 269 849

EA 018 378

AUTHOR Knoop, Robert; Common, Ronald W.
TITLE A Performance Appraisal System for School Principals.
PUB DATE May 85
NOTE 36p.; Paper presented at the Annual Meeting of the Canadian Society for the Study of Education (Montreal, Quebec, Canada, May 28-31, 1985).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Administrator Evaluation; Computer Oriented Programs; Data Interpretation; Elementary Secondary Education; *Evaluation Criteria; *Evaluation Methods; Evaluators; Foreign Countries; *Formative Evaluation; Interrater Reliability; Leadership Styles; Measurement Techniques; Observation; Performance; Performance Factors; *Principals; Tables (Data)
IDENTIFIERS Ontario

ABSTRACT

The Performance Review, Analysis, and Improvement System for Educators (PRAISE) is a formative evaluation instrument designed to improve the performance of school principals. The system appears to be reliable and valid and is flexible enough to accommodate the needs of a variety of schools. Sample items and categories of the instrument include problem solving and decision making; curriculum and program development; relations with teachers; relations to parents and community; and professional qualities and competencies. Selected categories and items of the inventory system for principal performance indicators include setting and achieving goals and objectives; directing and leading; managing change and innovations; and delegating and inviting participation. The inventory system is computerized to facilitate data processing. System features include the identification of strengths and weaknesses; a two-dimensional leadership profile; experiential exercises designed to analyze one's own actions and behavior in an objective manner; summaries for any segment of a population of educators; and training plans that show training needs by priority listings. PRAISE is an evolving system that, when compared to similar appraisal systems, shows great promise. Ten references are cited. (GJ)

 * Reproductions supplied by EDRS are the best that can be made *
 * from the original document. *

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

- ☒ This document has been reproduced as
received from the person or organization
originating it.
- ☐ Minor changes have been made to improve
reproduction quality.

• Points of view or opinions stated in this docu-
ment do not necessarily represent official NIE
position or policy.

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

B. Knoop

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

A PERFORMANCE APPRAISAL SYSTEM
FOR SCHOOL PRINCIPALS

Paper Presented at the Canadian Society for the
Study of Education
Montreal, May, 1985

Robert Knoop
Ronald W. Common
Brock University

ED269849

EA 018 378

A. INTRODUCTION

In appraising the performance of personnel, the literature generally differentiates between judgmental and developmental processes (Cummings & Schwab, 1973). The former is often termed summative evaluation, the latter formative (Howsam, 1973).

The PRAISE process (an acronym for Performance Review, Analysis & Improvement System for Educators) is formative; it was designed to give a person feedback on how he or she is perceived to perform at work, to bring out strengths and weaknesses, to guide the person to analyze the results, to work with them, to think about them, to overcome weaknesses, at least partially, and to reinforce strengths. In short, we designed PRAISE to improve performance.

Initially, some six years ago, we created an instrument to evaluate the performance of principals as a class project. About 40 Master of Education students then used this instrument to assess their own performance. The experience was not very satisfying or illuminating. Next, they decided to ask their supervisor, the superintendent, to rate their performance also. Now each student could compare the self ratings with the supervisor ratings. The process became much more interesting, more involving. Disagreements arose, objections to some of the ratings, which led to subsequent discussions with the supervisor. Then we decided to go one step further, to include peers, fellow principals with whom the principal-student had worked for years. Peers were more hesitant to participate, many requested anonymity by having the whole group, or family, of principals included as raters.

Two things happened. The process became cumbersome, and the results more difficult to analyze. For instance, a typical principal had to

manipulate one set of self ratings (I believe the instrument at that time contained 100 statement classified into eight categories), one set of supervisor ratings, and three to five sets of peer ratings. Mistakes were made in establishing mean scores, deviations; analysis was restricted because of the mass of numbers. Two or three of the student-principals added further to the appraisal process by asking their teachers to rate their own performance. Now they had four rater groups and up to 25 sets of ratings to contend with. Weeks passed before the data were analyzed. Most students had no energies left to engage in improvement, in thinking about the findings, contemplating them, making them fully conscious. Yet the data also provided insights, surprises, both positive and negative, resolves to "set perceptions straight", that were far deeper and of greater meaning than the solitary ratings we had started out with.

Of course, our difficulty led to computerization of the process. And because we started out fresh, we wanted a valid and reliable instrument. The Ontario Ministry of Education liked the idea of PRAISE and provided the funding that led to the development of the principal instrument that follows. In the last part of this paper, we will briefly describe the "Inventory System" which has since supplemented, if not supplanted, the the "fixed instrument" principle, and outline some system features of the PRAISE process.

B. FIXED INSTRUMENT

Initially, our intent was to design a unique, comprehensive performance appraisal instrument for school principals that is both reliable and valid. Recently, because of situational differences in goals, duties, responsibilities, job descriptions among school boards across Canada, we

decided to create a flexible system for selecting major job activities and specific job competencies that will be described in Part C of this paper.

Statement Selection

A government-sponsored search for representative competency statements for principals led to a series of research articles and books which were carefully perused and analyzed for statements that had been subjected to reliability and validity assessments (Knoop, 1979).

The search resulted in approximately 500 statements, that is, qualities, behaviours, duties, competencies, attitudes, responsibilities an effective principal should possess. This list of statements was mailed to a randomly selected sample of teachers (N = 50), principals (N = 40), and superintendent/directors (N = 20). Respondents were asked to rate the importance of each statement on a five-point scale, ranging from 1, extremely important, to 5, not important. The response was fairly high (50% for teachers, 54% for principals, and 40% for superintendents/ directors) even though the rating procedure took about 1 1/2 hours.

The responses were entered into a computer and rank-ordered by rater group and mean importance scores. The significance of the differences between rater groups were tested to assess the degree of perceptual role congruity of each statement; overall, particularly on statements deemed important, teachers, principals and superintendents did not disagree much on what an effective principal should be able to do, how he or she should behave, etc.

As a final step, four professors in education were asked to edit,

to detect duplicates and similarities in meaning of, the best 200 statements. Of these we retained, rather arbitrarily, 100 statements. Other studies have employed similar procedures (e.g. Chad, 1978; Florida State Project, 1976).

Factor Analysis

The instrument containing the 100 statements considered most important as indicators of behaviors, qualities, skills, knowledge, responsibilities and competencies of effective school principals was then given to 132 teachers who were asked to rate the performance of their principal on a "1" (Not at all characteristic of this principal) to "7" (Very characteristic of this principal) scale. The ratings were subsequently factor-analyzed.

Oblique rotations resulted in thirteen factor patterns. Loadings of .35 and above were considered adequate for inclusion in the next instrument. In this way, 81 items qualified. Loadings ranged from .36 ("Initiates, designs, and implements programs to meet the specific needs of the school") to .74 ("Is trusted and confided in by his/her teachers"); the median loading was .62. Eigenvalues exceeded 1.0.

The statements under each factor were then analyzed to detect similarities, preferably theory-based, to derive appropriate factor names. For instance, one cluster of statements pertained to problem solving and decision making. Items that could not be theoretically linked to the process of solving problems or making decisions were removed from that factor. To avoid proliferation of factors (categories), three factors (staffing, staff appraisal, and staff development) were combined into one category, another factor with only two statements was eliminated and several items with high loadings ($> .50$) that did not belong theoretically or

logically with other items under a particular factor, were combined into a factor entitled "Professional Qualities and Competencies". Ultimately this procedure resulted in 81 statements and 9 factors, or categories.

The 81-item instrument was then given to a sample of teachers (approximately 320 teachers enrolled in graduate studies at Brock University) who were asked to rate the performance of their present school principal. The response scale (from 1 to 7) was the same as in the first sample. Because of the collection procedure used---all professors teaching summer school were given sufficient instruments to distribute and collect at a specific time and date---a high percentage (82%) of the instruments were filled out and returned.

The responses were then again factor-analyzed. Many of the items loaded highly ($> .40$) on the same factors as in the first sample. For instance, items relating to Relations with Teachers, Students, Superintendents, Parents and Community, Problem Solving, Staffing, Curriculum and Program Development were distinctly separate from each other and from other factors. As in the previous sample, the items and factors for Staffing, Staff Appraisal, and Staff Development were put together under one heading.

Several of the nine items that make up the category Professional Qualities and Competencies had loadings that were distributed over several factors in excess of .35 even though the highest loading occurred on only one factor, the one we termed Professional Qualities. We retained these items because the findings "made sense". One would expect a statement such as "is ethical in all relationships" to load significantly on such factors as Relations with Teachers, Problem Solving, Relations with Superintendents in addition to a high loading on the category proper.

Two of the 81 items loaded below .40 and were eliminated from the item pool.

The 256 responses were also checked for internal consistency reliability. As shown in Table 1, alpha coefficients for all categories were extremely high (from .88 for Problem Solving and Decision Making to .98 for Relations with Superintendents).

Table 1
INTERNAL CONSISTENCY RELIABILITY ESTIMATES (ALPHA)

<u>Category</u>	<u>Items</u>	<u>Alpha</u>
Managing	10	.93
Problem Solving and Decision Making	9	.88
Staffing, Staff Appraisal and Staff Development	9	.92
Curriculum and Program Development	9	.96
Relations with Superintendents	9	.98
Relations with Teachers	9	.92
Relations with Students	9	.95
Relations to Parents and Community	8	.94
Professional Qualities and Competencies	9	.93
(All Items)	81	.99)

N = 256

Four groups from the overall sample were given the same instrument and the same instructions again at the end of a four-week course to test

the instrument for test-retest reliability. The results are given in Table 2. Estimates ranged from .59 for Relations to Parents and Community to .80 for Professional Qualities and Competencies, with an overall coefficient of .75 for the complete appraisal form.

Table 2
TEST-RETEST RELIABILITY ESTIMATES (PEARSON'S)

<u>Category</u>	<u>Items</u>	<u>Test-Retest Medians</u>
Managing	10	.74
Problem Solving and Decision Making	9	.69
Staffing, Staff Appraisal and Staff Development	9	.78
Curriculum and Program Development	9	.75
Relations with Superintendents	9	.77
Relations with Teachers	9	.68
Relations with Students	9	.64
Relations to Parents and Community	8	.59
Professional Qualities and Competencies	9	.80
(All Items	81	.75)

N = 102

The analyses discussed in the following section of this paper are based on a sample of 15 principals who volunteered to have their performance appraised by 36 superiors, 42 peers, and 119 subordinates. These principals also completed a self evaluation. This sample was smaller than we had anticipated. It seems that principals in the field are as yet not familiar with formative appraisals. Despite numerous invitations to participate, few principals were courageous or curious enough to find how

they are perceived to perform.

The means and standard deviations for the total sample are given in Table 3. They range from 5.40 (Problem Solving and Decision Making) to 6.03 (Professional Qualities). All means and standard deviations are fairly high and rather similar. The wide variation in N suggests that many raters did not have the knowledge to rate a principal on all items.

TABLE 3
MEANS AND STANDARD DEVIATIONS

	N	Mean	SD
Managing	89	5.70	.69
Problem Solving and Decision Making	154	5.40	1.04
Staffing, Staff Appraisal and Staff Development	77	5.43	.99
Curriculum and Program Development	94	5.42	1.05
Relations with Superintendents	74	5.73	.90
Relations with Teachers	161	5.40	1.20
Relations with Students	141	5.87	.85
Relations to Parents and Community	108	5.71	.86
Professional Qualities and Competencies	89	6.03	.71

The intercorrelations in Table 4 show that all categories were significantly ($p > .001$) correlated. The correlations range from .45 for Relations with Parents and Community, and Relations with Superintendents, to .84 for Curriculum and Program Development, and Staffing, Staff Appraisal and Staff Development. While one can no doubt derive interesting

TABLE 4
CATEGORY INTERCORRELATIONS

	1	2	3	4	5	6	7	8	9
(1) Managing	-								
(2) Problem Solving and Decision Making	.72	-							
(3) Staffing, Staff Appraisal and Staff Development	.71	.74	-						
(4) Curriculum and Program	.73	.76	.84	-					
(5) Relations with Superintendents	.49	.62	.46	.53	-				
(6) Relations with Teachers	.58	.80	.67	.70	.50	-			
(7) Relations with Students	.73	.77	.73	.76	.57	.79	-		
(8) Relations to Parents and Community	.62	.46	.61	.74	.45	.57	.71	-	
(9) Professional Qualities and Competencies	.73	.71	.72	.68	.66	.57	.65	.58	-

N's varied from 52 to 212

P>.001 for all intercorrelations

TABLE 5
INTERNAL CONSISTENCY RELIABILITY ESTIMATES (ALPHA)

Category	Self	Superint.	Peers	Teachers	Overall
Managing	.82 (10)	.32 (12)	.44 (12)	.77 (59)	.76 (89)
Problem Solving and Decision Making	.66 (9)	.89 (16)	.64 (21)	.90 (108)	.89 (154)
Staffing, Staff Appraisal and Staff Development	.51 (7)	.71 (8)	.78 (11)	.88 (51)	.86 (77)
Curriculum and Program Development	.58 (7)	.93 (10)	.87 (14)	.93 (63)	.92 (94)
Relations with Superintendents	.93 (7)	.86 (17)	.82 (17)	.82 (33)	.85 (74)
Relations with Teachers	.69 (9)	.90 (17)	.77 (22)	.92 (113)	.91 (161)
Relations with Students	.72 (9)	.70 (15)	.27 (19)	.87 (98)	.84 (141)
Relations to Parents and Community	.75 (7)	.89 (12)	.48 (18)	.87 (71)	.85 (108)
Professional Qualities and Competencies	.56 (7)	.81 (16)	.75 (19)	.83 (47)	.79 (89)

Note: The first number in each column is the alpha coefficient; the number in brackets is the number of cases (N). N varies because a listwise deletion procedure for each category was employed.

relationships and possible explanations for these relationships from these data, it has to suffice here to say that all categories were significantly correlated.

The high intercorrelations are in a way not surprising if one assumes that an "effective" principal is probably effective in a number of ways, while an ineffective principal performs generally poorly on various performance dimensions. The categories can therefore be considered useful separations of related dimensions.

Reliability

The internal consistency reliability estimates in Table 5 for the field sample are, overall, slightly lower than for the teacher sample. Teacher ratings of their principal generally had the highest reliability estimates, compared with self, superintendent, and peer ratings. For peers, the items relating to Managing (.44), Relations to Students (.27), and Relations to Parents and Community (.48) do not seem to measure the same thing consistently. For superintendents, Managing of the school received a low rating (.32), while the remaining estimates appear acceptable. Self ratings showed generally the lowest reliability coefficients with the exception of Relations to Superintendent (.93) and Managing (.82). In general, the reliability of the appraisal form for principals is fairly high. Indeed, when one considers the length and multi-dimensionality of the form, one could not have expected much higher reliability coefficients. That for a few tasks one or the other rater group did not seem to have the same knowledge or perception of how well a principal performed his or her duties and responsibilities is also not surprising. This finding suggests that to measure performance reliably we may want to leave out some items for certain rater groups. One can equally argue

though that raters could make an effort to become better informed. Is it not realistic to expect a superintendent, the direct supervisor of a principal, to know how well that principal manages his or her school?

Leniency Errors

In Table 6, the means and standard deviations by rater group and category are given. All mean ratings are above 5.0 which indicates that the four groups of raters thought highly of the participating principals. The findings suggest that the principals who appraised their own performance did not have an unrealistically high opinion of their own performance. The t-tests between self ratings and ratings of other groups show no significant differences along categories. Only one of the 27 possible combinations showed a difference: peers perceived principals as more effective in Curriculum and Program Development than principals did themselves. The results show therefore that the instrument is almost free from leniency errors.

Convergent Validity

Convergent validity compares the extent of agreement between two or more measures, or ratings, of the same trait, or category. Differences in ratings by the same group of raters show discrimination between traits. The procedure that estimates convergent validity using an analysis of variance technique is described by Kavanagh, MacKinney and Wolins (1971).

The results in Table 7 show that the instrument differentiates among subjects (principals), among categories (traits), and among raters. The significant subject effect is evidence of convergent validity: there is good differentiation among principals which can be attributed to the instrument used.

Table 6
Means, Standard Deviations and T-Tests
Between Self Ratings and Superior, Peer, and Subordinate Ratings

	<u>Self</u>		<u>Superintendents</u>			<u>Peers</u>			<u>Teachers</u>		
	M	SD	M	SD	t	M	SD	t	M	SD	t
Managing	5.68	.56	5.87	.39	.72	6.15	.34	1.88	5.58	.77	.41
Problem Solving and Decision Making	5.59	.57	5.27	.96	1.05	5.96	.50	1.67	5.29	1.13	1.39
Staffing, Staff Appraisal and Staff Development	5.41	.51	5.38	.62	.13	5.94	.78	1.77	5.33	1.11	.33
Curriculum and Program Development	5.14	.58	5.38	.99	.62	5.91	.59	2.86*	5.35	1.15	.79
Relations with Superintendents	5.47	1.14	5.36	1.01	.22	5.79	.82	.68	5.94	.78	1.05
Relations with Teachers	5.67	.56	5.41	1.09	.79	5.87	.67	.88	5.29	1.31	1.70
Relations with Students	5.89	.65	5.81	.67	.29	6.15	.41	1.11	5.82	.95	.27
Relations to Parents and Community	5.20	.86	5.93	.62	1.96	5.73	.48	1.54	5.71	.95	1.50
Professional Qualities and Competencies	5.95	.52	5.87	.68	.32	6.12	.73	.64	6.06	.75	.50
N	15		36			42			119		

* < .05

Unfortunately, due to the limitation of our computer, we were unable to calculate higher order interactions which could have given us estimates of discriminant validity and halo error. Instead we had to use other procedures.

Table 7
Analysis of Variance

Source of Variation	Sum of Squares	DF	Mean Square	F
Main Effects	229.42	19	12.07	16.23 ^{xxx}
Subjects (Principals)	162.77	8	20.35	27.35 ^{xxx}
Traits (Categories)	42.89	8	5.36	7.21 ^{xxx}
Raters	18.45	3	6.15	8.27 ^{xxx}

$p < .001$

Constant Errors

We followed a procedure used by Lee, Malone and Greco (1981) to determine constant errors. Means, standard deviations, skewness and kurtosis of the ratings for each category and for each and all rater groups are given in Table 8

Skewness is a measure of symmetry of a distribution about some central value of a variable. In a symmetrical distribution, mode, median and mean, all have the same value and skewness of zero. When the extended part of the distribution (the long tail) is to the left, as is the case for the combined rater scores (the midpoint of the rating scale, which went from 1 to 7, is 3.5), the skewness estimates are negative. Negative skewness

Table 8

Category	All Rater Groups				Self Ratings				Constant Errors				Superintendents				Peers			
									Teachers											
	M	SD	Kurtosis	Skewness	M	SD	Kurtosis	Skewness	M	SD	Kurtosis	Skewness	M	SD	Kurtosis	Skewness	M	SD	Kurtosis	Skewness
M	5.70	.69	.23	.87	5.68	.56	-1.18 ^a	-.28	5.58	.77	-.39	-.50	5.87	.39	-.50	-.64	6.15	.34	2.87 ^a	-1.53 ^a
PS & DM	5.40	1.04	-.48	-.59	5.59	.57	.49	.51	5.29	1.13	-.85	-.41	5.27	.96	-.85	-.29	5.96	.50	-.11	-.20
Staffing	5.43	.99	1.03 ^a	-.93	5.41	.51	-.90	.84	5.33	1.11	.47	-.84	6.11	.63	-.95	-.41	5.95	.78	2.46 ^a	-1.32
Curr.	5.42	1.05	.35	-.87	5.14	.58	.18	-.07	5.35	1.15	-.20	-.71	5.38	.99	3.52 ^a	-1.76 ^a	5.91	.59	-.18	-.58
R.SV	5.73	.90	-.13	-.69	5.47	1.14	2.15 ^a	-1.15 ^a	5.94	.78	-.78	-.49	5.36	1.01	-.94	-.40	5.79	.82	-.47	-.63
R.T.	5.40	1.20	-.27	-.77	5.67	.56	4.41 ^a	1.79 ^a	5.29	1.31	-.73	-.61	5.41	1.09	-.91	-.56	5.87	.67	-.25	-.41
R.St.	5.87	.85	.31	-.90	5.89	.65	-.59	-.10	5.82	.95	-.20	-.81	5.81	.67	-.20	-.39	6.15	.42	-.84	.13
R.Par.	5.71	.86	-.02	-.68	5.20	.86	-1.74 ^a	-.61	5.71	.95	-.37	-.64	5.93	.62	.54	-.86	5.73	.48	1.45 ^a	-.09
P.O.	6.03	.71	2.10 ^a	-1.18 ^a	5.95	.52	-.83	.53	6.06	.75	1.98 ^a	-1.21 ^a	5.87	.68	-1.21 ^a	-.04	6.12	.73	8.58 ^a	-2.49 ^a

p < .01

indicates that the median is larger than the mean. Whenever the skewness index is larger than one, the distribution is considered to be highly skewed (Hill and Kerber, 1967).

The data in Table 8 show that the ratings are generally not very skewed even though most coefficients are negative. One of the nine categories (Professional Qualities) for the combined raters shows significant skewness; this seems mainly due to peer ratings ($-2.49, p < .01$) and teacher ratings; ($-1.21, p < .01$). Skewness is also indicated for two categories of self ratings (Relations with teachers and superintendent), for one category of superintendent ratings (Curriculum and Program Development), one category of teacher ratings (Professional Qualities), and three categories of peer ratings (Managing, Staffing, and Professional Qualities).

An examination of the standard deviations for each performance category and rater group suggests that the raters generally did differentiate in their perceptions of performance effectiveness to a reasonable, if varying, degree. As the reader knows, standard deviation is by far the most commonly used indicator of degree of dispersion and the most dependable source of estimates of variability in the total population.

The smallest standard deviations occurred for self ratings and peer ratings, among rater groups, and for Professional Qualities and Managing, among categories. Of course, standard deviations are influenced by the size and nature of the sample: with an increase in sample size and a decrease in volunteering principals, variability and dispersion should increase and the means should drop.

Kurtosis estimates the degree of steepness of the middle part of the distribution of ratings. A normal distribution is mesokurtic which means neither very peaked nor very flat across the top. Those more peaked than

Table 9

Intercorrelations of Performance Ratings

Medians & Ranges for Categories

Category	All Rater Groups		Self Rating		Teachers		Superintendent		Peers	
	Median	Range	Median	Range	Median	Range	Median	Range	Median	Range
M	.22	.10 to .39	.35	.14 to .80	.30	.02 to .41	.08	.42 to .80	.05	.48 to
PS	.41	.31 to .78	.31	.12 to .69	.50	.33 to .77	.47	.10 to .88	.20	.17 to
ST	.32	.14 to .59	.23	.31 to .92	.39	.15 to .61	.27	.23 to .72	.18	.25 to
Cur	.51	.21 to .76	.30	.46 to .84	.57	.22 to .80	.43	.04 to .79	.37	.04 to
RSU	.40	.26 to .72	.65	.24 to .95	.45	.23 to .87	.47	.13 to .83	.41	.11 to
RT	.60	.11 to .82	.42	.29 to .84	.65	.09 to .82	.74	.05 to .92	.48	.24 to
RST	.52	.14 to .74	.50	.27 to .88	.51	.19 to .75	.40	.22 to .84	.24	.26 to
RP	.38	.09 to .66	.35	.48 to .97	.41	.14 to .68	.42	.12 to .87	.30	.47 to
PQ	.37	.03 to .63	.34	.56 to .81	.34	.04 to .62	.45	.46 to .85	.16	.28 to

normal are termed leptokurtic. The Kurtosis estimates in Table 8 indicate that only two of the nine overall category ratings show evidence of Leptokurtosis, a symptom of central tendency errors in ratings. For the individual rater groups, these errors appear more frequently for self and peer ratings (4 of 9 categories) than for teacher (1 of 9) and superintendent (2 of 9 categories) ratings. Leptokurtosis is therefore evident to some degree.

Because of the nature of the sample, volunteers, some indication of constant errors is not surprising. These principals may well have been those who were confident that their ratings would be very positive.

Halo Errors

Halo occurs when raters do not differentiate among distinct items, or categories, in their evaluation of a rater, but instead base their evaluation on some global or overall judgment.

The procedure employed is based on an analysis of median inter-correlations among performance categories (Lee, Malone and Greco, 1981), both for combined and individual rater groups (Table 9). The median correlation coefficients range from .22 to .60 for all rater groups, .23 to .65 for self ratings, .30 to .65 for teacher ratings, .08 to .74 for superintendent ratings, and .05 to .48 for peer ratings.

These values indicate that halo error exists for some categories and rater groups. On closer inspection, however, and taking into account the range of correlation coefficients, the data show a fair degree of discrimination among items within categories. Indeed, many ranges include negative correlations among items which indicates that a principal can for instance, "work with a superintendent in planning school programs and activities" because he or she is obliged to do so, without having

"confidence and trust in his/her superintendent".

One can argue that the data show that halo effects are minimal, as evidenced by the wide range of coefficients, and that the instrument enables raters to evaluate the performance of principals fairly accurately on a variety of distinct items pertaining to one category.

Conclusion

The instrument we developed to assess the effectiveness of principal performance appears to be reliable and valid. As is the case with all instruments, more work is needed to test other aspects of validity, to establish a norms base, to relate this appraisal form to other field instruments, and to apply it to non-volunteers.

A sample of the items and categories that make up the fixed instrument is given in Table 10.

C. FLEXIBLE INSTRUMENT

Numerous principals, particularly those from small Ontario boards and from outside the Province of Ontario, complained that not all statements of the fixed instrument just described were applicable to their specific situation. For instance, not all were involved in staffing, many did not have a mechanism for staff evaluation, let alone a systematic, long-range plan for staff development. Yet some of the boards managed by objectives, with a different set of objectives in each new school year, which were not reflected in the fixed instrument.

To overcome this difficulty, we decided to create an item bank of statements and categories. We carefully accumulated once again performance indicators from the literature and added those to our fixed instrument.

Table 10

Sample Items and Categories of the Performance
Appraisal Instrument for Principals

PROBLEM SOLVING AND DECISION MAKING

- ___ 11. Is unable to identify real, important, and relevant problems.
- ___ 12. Systematically analyzes a problem, making use of all available information.
- ___ 13. Generates an appropriate number of feasible solution alternatives.
- ___ 14. Does not anticipate the consequences of, and reactions to, each alternative in reaching a decision.

CURRICULUM AND PROGRAM DEVELOPMENT

- ___ 29. Has success in utilizing the talents and ideas of faculty and consultants in improving curriculum.
- ___ 30. Initiates, designs, and implements programs to meet the specific needs of the school.
- ___ 31. Determines curriculum areas needing modification.
- ___ 32. Assists staff members to bring about continuity of students' learning experiences, both within and between grades.
- ___ 33. Has not developed in co-operation with staff a written statement of school philosophy and objectives.

RELATIONS WITH TEACHERS

- ___ 45. Displays behaviour which enhances teachers' feelings of personal worth and esteem.
- ___ 46. Has confidence and trust in his/her teachers.
- ___ 47. Is trusted and confided in by his/her teachers.
- ___ 48. Does not have the respect of his/her staff.
- ___ 49. Has good communications with teachers.

RELATIONS TO PARENTS AND COMMUNITY

- ___ 62. Disseminates information about school activities and programs to community.
- ___ 63. Does not have a planned program for school-community relations to further the understanding and support of the school program.
- ___ 64. Maintains an awareness of community factors which affect the school program.

PROFESSIONAL QUALITIES AND COMPETENCIES

- ___ 70. Has a thorough understanding of what is expected of his/her role as principal.
- ___ 71. Is ethical in his relationships.
- ___ 72. Shows enthusiasm and drive in his/her work.
- ___ 73. Accepts responsibility of his/her actions without blaming others.

Table 11

Selected Categories and Items of the Inventory System
for Principal Performance Indicators

CATEGORY: Setting and Achieving Goals and Objectives

1. Aims to have the best school in the district.
2. Generally achieves objectives for his/her schools.
3. Has established goals and objectives to give the school, and its teachers and students, direction and purpose.
4. Holds high performance goals for his/her school.

CATEGORY: Directing and Leading

1. Allows momentary chaos.
2. Gets teachers from where they are to where they have not been.
3. Informs staff of the duties and responsibilities of the leadership team.
4. Inspires teachers to become managers of learning.
5. Shapes the school.
6. Helps teachers form and attain goals.

CATEGORY: Motivating

1. Attempts to provide a school environment (psychological and physical) that is conducive to the learning process.
2. Creates an exciting school environment through personal attention.
3. Creates and promotes school values.
4. Knows how to get teachers to do what has to be done.

CATEGORY: Planning

1. Knows how his/her school can get from where it is now to where it wants to be.
4. When planning, relates student/parent needs to school goals.
5. Creates an effective-master timetable for the school.

CATEGORY: Organizing

1. Has arranged for efficient secretarial and office services.
2. Has organized the school to produce good interrelationships among teachers.
3. Has set up a good system for keeping teachers and students informed.
4. Integrates: brings about unity of effort among teachers.

CATEGORY: Managing Change and Innovations

1. Encourages risk-taking and experimentation.
2. Actively involves teachers to be affected by a change program in the planning and implementation process.
3. Actively seeks ideas from teachers.
4. Demonstrates an open attitude to teacher-initiated change.

CATEGORY: Handling Conflict

1. Knows how to deal with personality conflicts between staff members.
2. Encourages "give-and-take" discussions with teachers.
3. Focuses feedback on specific behaviour rather than on the person.
4. In conflict situations, skillfully questions others' basic assumptions and convictions.

CATEGORY: Delegating and Inviting Participation

1. Consults teachers before acting.
2. Knows that for some people participation does not motivate.
3. Knows that participation reduces problems of absenteeism and turnover, complaints and grievances.
4. Knows that participation gives most people a sense of ownership, a feeling of competence and self esteem.

CATEGORY: Staffing

1. Has the ability to select outstanding teachers for the school.
2. Knows that school effectiveness depends on staffing decisions.

Many of these additional statements were the theme of whole articles; others recurred regularly in all types of journals, refereed and non-refereed. Whenever possible, we searched for validity and reliability data, for factor loadings and headings. Table 11 contains some of the items and categories of the Inventory of Performance indicators for principals not shown in the fixed instrument.

The Inventory System is computerized to facilitate data processing. If, for example, principals of a given board want to create "their own" appraisal form, each participant or representative participant, can rate the importance of each item to the role of principal in that board. They can also add items or categories not contained in the Inventory, a choice that is rarely exercised because of the relative completeness of the inventory. The principals can also give a set of Inventories to their supervisors, superintendents, and/or to their subordinates, teachers.

All responses are fed into the computer. On the last page of the inventory each participant is asked for the exact number of items, or for a range, the final appraisal form should consist of. The computer then computes means for each item, rank-orders and categorizes the responses and prints out the form. Or, if selected, the board can get feedback on the degree of congruency for each item; that is, the mean scores for each rater group and the overall means are shown separately. This latter procedure can lead to increased communication among educators within a board on the role of a principal. The former procedure results in an appraisal form most participants can subscribe to.

The problems associated with a flexible, versus a fixed, instrument relate to the age-old dilemma of reliability and validity. This problem we have not yet solved, but we are experimenting with potential solutions.

Once a board has agreed on an appraisal form, it is quite possible to provide, on command, outputs on internal consistency reliability estimates, test-retest reliability and various types of validity. The ANOVA procedure developed by Kavanagh, MacKinney and Wolins (1971) would be one suitable procedure to determine the degree of convergent and discriminant validity, as well as the effect of method bias.

Similarly, a randomly selected single item can be duplicated on each instrument and printed by the computer to provide test-retest estimates.

In other words, we want to create a system that can either give us a report on the reliability and validity of a particular instrument used, or to program the computer to drop automatically items that do not meet predetermined standards.

Concerns of reliability and validity are of course of utmost importance in performance evaluation. Without such evidence the system highlighted on the next few pages may lack the impact it was designed to have on performance effectiveness.

D. SYSTEM FEATURES

It is beyond the scope of this paper to describe all ramifications of PRAISE. Here we want to outline only a few features that we believe should be present in any good formative appraisal system.

Weaknesses in performance are identified based on the lowest mean scores by statements and categories. The means compare self, supervisor, peer and subordinate ratings with family of schools, board and provincial (computer) norms. In other words, the multiple comparisons allow a participant to assess how pronounced a weakness really is.

Strengths are based on the highest mean scores, again by statements and categories, separate for each rater group. The ratings are then compared against the three norms.

Rater Disagreements show the largest perceptual incongruencies between self ratings and ratings of each group of raters who participated in the appraisal.

Rater Agreements compares self ratings with each of the other rater groups to identify those performance activities on which there is the highest perceptual congruency.

Leadership Profile is two-dimensional and shows the extent of people and task orientation, along the lines of other two-dimensional theories (Ohio State research, Michigan Studies, Blake and Mouton, Fiedler, and others) for each rater group. Items were originally based on modified statements of the LBDQ, Form 12, but are now measured globally ("This principal is basically a people-oriented leader").

A second "picture" of the leadership profile shows the combined ratings of all those who rated a principal against the combined perception of what the position in a particular school requires. The degree of overlap of the two areas, "is" and "should be", shows the match, the fit of leader and situation.

A third and fourth window of the profile compare the individual's leadership style against styles of other principals in the board and the computer (provincial norm) and against the preferred style within board and province.

Experiential Exercises are data-related and designed to open up insights, to show connections, to bring to awareness what was unconscious, to invite speculation, to analyze one's own actions and behaviour in an objective

way. The exercises are based on the raw findings, on the analysis part of the data, and on additional computer-guided examination of specific aspects of the performance ratings.

Each exercise can be printed separately and can in this way provide the basis for professional development activities (see also Training Plans). Overviews is a term we selected to denote short, 2,000 to 5,000 word, "state-of-the-art" reports for the various categories. The Overviews are an attempt to bridge the ever-widening gap between theorists and practioners by presenting in simple language what we presently know about a topic. Participants receive overviews only for pronounced performance weaknesses.

Summaries are menu-based and can be selected for any segment of a population of educators, from summaries for an individual principal to family of schools to board-wide summaries. To preserve anonymity of individuals two precautions were added: to use only I.D. numbers, and to prevent printing of summaries for groups of principals smaller than 3. The summaries also include comparison data with other groups; for instance, a district superintendent can compare the performance of principals in his/her family of schools against the performance of other families within the board, and against the computer (which could become the provincial) norm.

Training Plans show training needs by priority listings. For example, for one board the greatest training priority for elementary school principals may be "Setting and Achieving Goals and Objectives".

Specifically

the plan may show that "joint goal setting with teachers", "setting goals that are feasible and attainable", and "setting up policies that relate

to the attainment of objectives" need improvement.

Training plans are also menu-based and allow any combination of variables to be selected.

* * *

PRAISE is an evolving system for assessing and improving personnel performance. We have attempted in this paper to give an overview of the system and to describe one aspect of it, the development of a fixed instrument to appraise the performance of school principals, in some detail. Much work needs to be done to develop PRAISE further. Yet in comparison to similar appraisal systems, particularly in industry, the PRAISE process shows great promise.

References

- Blake, R., & Mouton, J. The Managerial Grid, Houston, Gulf Publishing, 1964.
- Chad, E. "Teacher Assessments of Principal Performance", Utah University, CCBC Notebook, 1978, 7, 4-21.
- Cummings, L., & Schwab, D. Performance in Organizations: Determinants & Appraisal, Glenview, Ill., Scott, Foresman, 1973.
- Fiedler, F., A Theory of Leadership Effectiveness, New York, McGraw-Hill, 1967.
- Exemplary Competency-Based Vocational Teacher Education Project: Career Education Component. Professional Competencies and Assessment Criteria for implementing the Career Education Concept in Vocational Education Programs in Florida. Final Report of Competency and Criteria Identification Phase of Career Education Component. Program of Vocational Education, Florida State University, Tallahassee, Florida, June 1976.
- Howsam, R. "Current Issues in Evaluation", National Elementary Principal, 1973, 52, 12-17.
- Kavanagh, M., MacKinney, A., & Wolins, L. "Issues in Managerial Performance: Multitrait-Multimethod Analysis of Ratings", Psychological Bulletin, 1971, 75, 34-49.
- Knoop, R. An Inventory of Competency Statements for Teachers, Principals, and Superintendents, Ontario Ministry of Education, 1979.
- Lee, R., Malone, M., & Greco, S. "Multitrait-Multimethod Analysis of Performance Ratings of Law Enforcement Personnel", Journal of Applied Psychology, 1981, 66, 625-632.
- Stogdill, R. Handbook of Leadership, New York, Free Press, 1974.